GENETIC AND GENETIC-ENVIRONMENTAL COMPONENTS OF GENERATION MEANS ORIGINATING FROM VARIETIES IN HARDY-WEINBERG EQUILIBRIUM

by

R. R. Corbeil

BU-544-M

January, 1975

Abstract

A general model of genetic effects is extended to include all first order epistatic effects in t fixed or random environments. Component estimation procedures are indicated and problems of practical concern discussed.

Introduction

The evaluation of genetic-environment (g-e) interactions in plant populations derived from varieties which are not pure breeding gives rise to a number of recognized problems, the most fundamental being identification of genotypic classes. In applications, the problem is circumvented by equating variety to genotype as in variety trial studies or by the regression of varieties on an environmental index [e.g., see Eberhart and Russell (1966), and Knight (1973)]. Neither approach, however, is amenable to a lucid genetic interpretation of the non-additive environmental effect on genotype. A more enlightening alternative is to consider the g-e interactions in terms of the parameters of a genetic model. This approach has received much attention, possibly originating with the work of Fisher and Mackenzie (1923). For an overview of the historical and recent developments in analyzing g-e interactions the reader is referred to Freeman (1973).

Paper No. BU-544-M in the Biometrics Unit Mimeo Series, Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14853.

One serious difficulty associated with relating g-e interactions to genetic parameters stems from the absolute dependence of the procedure on adequate modeling of genetic populations. A general genetical model of the type developed by Eberhart and Gardner (1966) helps alleviate the difficulty. The purpose of this paper is to extend the usefulness of the Eberhart-Gardner model by redirecting the objectives from prediction of the response of 3 and 4-way crosses among varieties in a diallel experiment to that of extracting genetic components in 1 and 2-way crosses to study g-e interactions. The present objective permits the identification of many first order inter-allelic interactions in crosses between varieties which, although not pure breeding, are in a Hardy-Weinberg equilibrium.

The Genetic Model

A model allowing for variety crosses in which pure lines are not a requirement must take into account the multiplicity of alleles which can exist in the varieties under study [Kempthorne (1957)]. Let p_{ij_ik} be the frequency of the gene A_{ij_i}, that is, the frequency of the j_i^{th} allele of the i^{th} locus in variety k. Given that the k^{th} variety has the genotypic array

$$\prod_{i=1}^{n} \left(\sum_{j_i=1}^{m_i} p_{ij_ik} A_{ij_i} \right)^2$$

and yields the gametic array

$$\prod_{i}^{n} \left(\sum_{j_i}^{m_i} p_{ij_ik} A_{ij_i} \right),$$

where n is the number of loci involved in the trait of interest and m_i is the number of alleles at the i^{th} locus, then this variety is said to have mean

(or entry) Y_k. Should these same conditions hold for some other variety k', say, with mean $Y_{k'}$, then the genotypic array of the cross between the k^{th} and k'^{th} varieties will be

$$\prod_i^n \left[\left(\sum_{j_i}^{m_i} p_{ij_ik} A_{ij_i}\right)\left(\sum_{j_i}^{m_i} p_{ij_ik'} A_{ij_i}\right)\right]$$

and have mean $Y_{kk'}$. It is possible to derive further generation means by restricting mating in the k^{th} variety to a series of selfing generations. The gametic array to be produced beyond the first selfed generation will not be

$$\prod_i^n \left[\tfrac{1}{2}\sum_{j_i}^{m_i} (p_{ij_ik} + p_{ij_ik'}) A_{ij_i}\right]$$

unless independence is assumed between A_{ij_i}'s. The generation mean for the selfing series may be represented by $Y_k^{s^g}$, where g is the number of selfed generations.

Basic Mendelian considerations allow for the derivation of other generation means. Let $Y_{kk'}^{r}$ be the mean of a cross between varieties k and k' which is further crossed by a random association of its gametes. Likewise, let $Y_{kk'}^{s^g}$ be the mean of a generation derived in g selfing generation steps. Finally, backcross generations and selfing series of backcrosses can be represented by their means as $Y_{kk'\cdot k}$ and $Y_{kk'\cdot k}^{s^g}$, respectively. These generation means are listed in the first column of Table 1 and represent the results of typical 1 and 2-way crosses among an arbitrary number of varieties.

The appropriate model for each generation mean is a function of the following primary genetic effects:

α_{ij_i} is the additive contribution of gene A_{ij_i} to the response of the trait of interest,

$\delta_{ij_iij_i'}$, for $j_i' \neq j_i$, is the dominance deviation for a locus of the type $A_{ij_i}A_{ij_i'}$,

$\alpha\alpha_{ij_ii'j_i'}$, $\alpha\delta_{ij_ii'j_i'}$, and $\delta\delta_{ij_iij_i'i'j_i'i'j_i'}$, for $i' \neq i$ and $j' \neq j$, are the additive × additive, additive × dominance, and dominance × dominance inter-allelic interactions, respectively.

Except in simple digenic crosses the primary genetic effects usually cannot be directly identified. Therefore, estimable functions of the primary effects must be devised for convenience and ease of interpretation. First consider the following definitions:

$$a_k = 2\sum_i \sum_{j_i} p_{ij_ik}\alpha_{ij_i},$$

$$d_k = 2\sum_i \sum_{j_i} \sum_{j_i'>j_i} p_{ij_ik}p_{ij_i'k}\delta_{ij_iij_i'},$$

$$h_{kk'} = \sum_i \sum_{j_i} \sum_{j_i'>j_i} (p_{ij_ik} - p_{ij_ik'})(p_{ij_i'k'} - p_{ij_i'k})\delta_{ij_iij_i'},$$

$$aa_k = 4\sum_i \sum_{i'>i} \sum_{j_i} \sum_{j_i'} p_{ij_ik}p_{i'j_i'k}\alpha\alpha_{ij_ii'j_i'},$$

$$aa_{kk'} = \sum_i \sum_{i'>i} \sum_{j_i} \sum_{j_i'} (p_{ij_ik} - p_{ij_ik'})(p_{i'j_i'k'} - p_{i'j_ik})\alpha\alpha_{ij_ii'j_i'},$$

$$ad_k = 4\sum_i \sum_{i'>i} \sum_{j_i} \sum_{j_i'} \sum_{j_i'>j_i'} p_{ij_ik}p_{i'j_ik}p_{i'j_i'k}\alpha\delta_{ij_ii'j_ii'j_i'},$$

$$ad_{kk'} = \sum_{i}\sum_{i'>i}\sum_{j_i}\sum_{j_{i'}}\sum_{j'_i,>j_{i'}} (p_{ij'_i,k'} - p_{i'j'_i,k})$$

$$\cdot (p_{ij_ik}p_{i'j_i,k} - p_{ij_ik'}p_{i'j_i,k'})\alpha\delta_{ij_ii'j_{i'},i'j'_i},$$

$$ad_{lkk'} = \sum_{i}\sum_{i'>i}\sum_{j_i}\sum_{j_{i'}}\sum_{j'_i,>j_{i'}} (p_{ij_ik'} - p_{ij_ik})$$

$$\cdot (p_{i'j_i,k}p_{i'j'_i,k} - p_{i'j_i,k'}p_{i'j'_i,k'})\alpha\delta_{ij_ii'j_{i'},i'j'_i},$$

(For definitions of da_k, $da_{kk'}$, and $da_{lkk'}$ reverse i and i' and replace $\alpha\delta$ with $\delta\alpha$ in the three definitions immediately above.)

$$dd_k = 4\sum_{i}\sum_{i'>i}\sum_{j_i}\sum_{j'_i>j_i}\sum_{j_{i'}}\sum_{j'_i,>j_{i'}} p_{ij_ik}p_{ij'_ik}p_{i'j_i,k}p_{i'j'_i,k}\delta\delta_{ij_iij'_ii'j_{i'},i'j'_i},$$

$$dd_{kk'} = \sum_{i}\sum_{i'>i}\sum_{j_i}\sum_{j'_i>j_i}\sum_{j_{i'}}\sum_{j'_i,\neq j_{i'}} (p_{ij_ik}p_{i'j_i,k} - p_{ij_ik'}p_{i'j_i,k'})$$

$$\cdot (p_{ij'_ik'}p_{i'j'_i,k'} - p_{ij'_ik}p_{i'j'_i,k})\delta\delta_{ij_iij'_ii'j_{i'},i'j'_i},$$

$$dd_{lkk'} = \sum_{i}\sum_{i'>i}\sum_{j_i}\sum_{j'_i>j_i}\sum_{j_{i'}}\sum_{j'_i,>j_{i'}} (p_{ij_ik}p_{ij'_ik} - p_{ij_ik'}p_{ij'_ik'})$$

$$\cdot (p_{i'j_i,k'}p_{i'j_i,k'} - p_{i'j_i,k}p_{i'j'_i,k})\delta\delta_{ij_iij'_ii'j_{i'},i'j'_i},$$

$$dd_{2kk'} = \sum_{i}\sum_{i'>i}\sum_{j_i}\sum_{j'_i>j_i}\sum_{j_{i'}}\sum_{j'_i,\neq j_{i'}} (p_{i'j_i,k'} - p_{i'j_i,k})$$

$$\cdot (p_{ij_ik}p_{ij'_ik}p_{i'j'_i,k} - p_{ij_ik'}p_{ij'_ik'}p_{i'j'_i,k'})\delta\delta_{ij_iij'_ii'j_{i'},i'j'_i},$$

and

$$dd_{3kk'} = \sum_{i}\sum_{i'>i}\sum_{j_i}\sum_{j'_i>j_i}\sum_{j_{i'}}\sum_{j'_i,\neq j_{i'}} (p_{ij_ik'} - p_{ij_ik})$$

$$\cdot (p_{ij'_ik}p_{i'j_i,k}p_{i'j'_i,k} - p_{ij'_ik'}p_{i'j_i,k'}p_{i'j'_i,k'})\delta\delta_{ij_iij'_ii'j_{i'},i'j'_i}.$$

The above functions have unambiguous genetic meaning. The first five are given in Eberhart and Gardner (1966) where the meaning and partitioning of $h_{kk'}$, the heterosis parameter, is discussed at length. Complete confounding reduces the identifiable set of parameters to 12. These are given on the top line of Table 1. The sum of the additive gene effects, a_k, is completely confounded with intra-variety additive x additive gene interactions, aa_k. Therefore, a new parameter is defined as $a_k^* = a_k + aa_k$. The identical situation exists between the sum of the dominance deviations and the intra-variety additive x dominance interactions: $d_k^* = d_k + ad_k + da_k$. Confounding effects also involve the inter-variety parameter $h_{kk'}$ with $ad_{kk'}$ and $da_{kk'}$, their sum being designated $h_{kk'}^*$. Finally, $dd_{2kk'} + dd_{3kk'}$ cannot be separated and are identified as $dd_{kk'}^*$.

The results are embodied in Table 1. The minimum experiment required consists of at least 12 generations for the simultaneous estimation of the given parametric functions. The simplest sub-model, the so-called additive-dominance model, includes a_k^*, d_k^*, and $h_{kk'}^*$, the main effects, which can be interpreted only in the absence of the epistatic effects aa_k, ad_k, and da_k. However, it seems reasonable in a loose sense to believe $aa_k = ad_k = da_k = 0$ when other parameters such as $aa_{kk'}$, $ad_{kk'}$, and $da_{kk'}$ are zero. One word of caution, however, $aa_{kk'}$, $ad_{kk'}$, and $da_{kk'}$ can on a rare occasion be zero without $\alpha\alpha$, $\alpha\delta$, and $\delta\alpha$ being zero.

In the special case of pure breeding varieties, only one allele is present at each locus, hence the expectation for a variety becomes $Y_k = Y_k^{ss} = a_k^*$. This is because $p_{ij_1k} = 1$ while $p_{ij_1'k} = 0$ so d_k and dd_k do not appear in the model equations. For the same reason, other parameters are excluded from the model as well; these are: ad_k, $ad_{1kk'}$, da_k, $da_{1kk'}$, $dd_{1kk'}$, $dd_{2kk'}$, and $dd_{3kk'}$. This leaves the model with the parameters a_k, $a_{k'}$, $h_{kk'}$, aa_k, $aa_{kk'}$, $ad_{kk'}$, $da_{kk'}$, and $dd_{kk'}$ or a_k^*, $a_{k'}^*$, $h_{kk'}^*$, $aa_{kk'}$, and $dd_{kk'}$, which are similar to those of Hayman (1954).

Table 1: The General Genetic Model for Mean Components

Generation Mean	a^*_k	$a^*_{k'}$	d^*_k	$d^*_{k'}$	dd_k	$dd_{k'}$	$h^*_{kk'}$	$aa_{kk'}$	$ad^*_{kk'}$	$dd_{kk'}$	$dd_{1kk'}$	$dd^*_{kk'}$
Y_k	1		1		1							
$Y_{k'}$		1		1		1						
Y^s_k	1		$\frac{1}{2}$		$\frac{1}{4}$							
$Y^s_{k'}$		1		$\frac{1}{2}$		$\frac{1}{4}$						
$Y^{s^g}_k$	1		$\left(\frac{1}{2}\right)^g$		$\left(\frac{1}{4}\right)^g$							
$Y^{s^g}_{k'}$		1		$\left(\frac{1}{2}\right)^g$		$\left(\frac{1}{4}\right)^g$						
$Y_{kk'}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	1		1		
$Y^s_{kk'}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{2}$	1		$\frac{1}{4}$		
$Y^{s^g}_{kk'}$	$\frac{1}{2}$	$\frac{1}{2}$	$\left(\frac{1}{2}\right)^{g+1}$	$\left(\frac{1}{2}\right)^{g+1}$	$\left(\frac{1}{2}\right)^{2g+1}$	$\left(\frac{1}{2}\right)^{2g+1}$	$\left(\frac{1}{2}\right)^g$	1		$\left(\frac{1}{4}\right)^g$		
$Y^r_{kk'}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$\frac{1}{2}\left(Y_{kk'\cdot k} + Y_{kk'\cdot k'}\right)$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$
$\frac{1}{2}\left(Y^s_{kk'\cdot k} + Y^s_{kk'\cdot k'}\right)$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{16}$		$\frac{1}{16}$
$\frac{1}{2}\left(Y^{s^g}_{kk'\cdot k} + Y^{s^g}_{kk'\cdot k'}\right)$	$\frac{1}{2}$	$\frac{1}{2}$	$\left(\frac{1}{2}\right)^{g+1}$	$\left(\frac{1}{2}\right)^{g+1}$	$\left(\frac{1}{2}\right)^{2g+1}$	$\left(\frac{1}{2}\right)^{2g+1}$	$\left(\frac{1}{2}\right)^{g+1}$	$\frac{3}{4}$	$\left(\frac{1}{2}\right)^{g+2}$	$\left(\frac{1}{4}\right)^{g+1}$		$\left(\frac{1}{4}\right)^{g+1}$

See text for explanation of symbols.

Genetic-Environmental Interactions

If each of the generations is grown in t environments, the ℓ^{th} environment contributes ϵ_ℓ with $\sum_\ell^t \epsilon_\ell = 0$ to the generation mean with the vector of interaction parameters being $(\omega_{a_k^*\ell}, \omega_{a_k^*, \ell}, \omega_{d_k^*\ell}, \cdots, \omega_{dd_{kk}^*, \ell}) = \{\omega_{q\ell}\}$ where the subscript, q, of each element of the vector is taken from the top line of Table 1.

Each $\omega_{q\ell}$ represents a non-additive contribution to the mean of a generation such that $\sum_\ell^t \omega_{q\ell} = 0$ for each q. The proper coefficient of $\omega_{q\ell}$ is the product of the coefficients of ϵ_ℓ and of q where $q = a_k^*, a_k^{*\prime}, \cdots, dd_{kk'}^*$, as before. As an example, the extended model for $Y_{kk'}$ in the ℓ^{th} environment is $Y_{kk'\ell} = \frac{1}{2}a_k^* + \frac{1}{2}a_{k'}^* + \frac{1}{2}d_k^* + \frac{1}{2}d_{k'}^* + \frac{1}{2}dd_k + \frac{1}{2}dd_{k'} + h_{kk'}^* + aa_{kk'} + dd_{kk'} + \epsilon_\ell + \frac{1}{2}\omega_{a_k^*\ell} + \frac{1}{2}\omega_{a_{k'}^*, \ell} + \frac{1}{2}\omega_{d_k^*\ell} + \frac{1}{2}\omega_{d_{k'}^*, \ell} + \frac{1}{2}\omega_{dd_k\ell} + \frac{1}{2}\omega_{dd_{k'}, \ell} + \omega_{h_{kk'}^*, \ell} + \omega_{aa_{kk'}, \ell} + \omega_{dd_{kk'}, \ell}$.

Estimation

The genetic parameters are unknown fixed values which need to be estimated from the observed generation means. The extended model, that is the model with environmental factors, may be of two kinds: (A) fixed, if the environmental factors are a chosen set of environments, and (B) mixed, if the environments are a random sample of environments. Under the fixed model a least squares or weighted least squares procedure can be used in a straightforward manner and most multiple regression computer routines will suffice. Let $\underset{\sim}{X}$ be the matrix of coefficients associated with the parameters of the extended model and $\underset{\sim}{y}$ be the vector of observed generation means in all t environments. Then if $\underset{\sim}{\beta}$ represents the vector of parameters included in the model equations for generation means, the weighted least squares solution for $\hat{\underset{\sim}{\beta}}$, say, is

$$\hat{\underset{\sim}{\beta}} = (\underset{\sim}{X}'\underset{\sim}{D}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{D}^{-1}\underset{\sim}{y}$$

where $\underset{\sim}{D}$ is a diagonal matrix of variances each being associated with an observed

mean. A test for the goodness of fit of the model is $\underset{\sim}{y}'\underset{\sim}{D}^{-1}\underset{\sim}{y} - \underset{\sim}{y}'\underset{\sim}{D}^{-1}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{D}^{-1}\underset{\sim}{X})\underset{\sim}{X}'\underset{\sim}{D}^{-1}\underset{\sim}{y}$ which is distributed as χ^2 with the number of degrees of freedom equal to the number of entries (means) less the number of parameters estimated. In the ordinary least squares procedure $\underset{\sim}{D} = \underset{\sim}{I}\sigma^2$ and, therefore, an estimate of σ^2 is obtained as a residual in the analysis of variance table. It follows immediately from least squares theory that

$$\widehat{\text{var}}(\underset{\sim}{\hat{\beta}}) = (\underset{\sim}{X}'\underset{\sim}{D}^{-1}\underset{\sim}{X})^{-1} \quad \text{or} \quad (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\hat{\sigma}^2$$

the $(\underset{\sim}{X}'\underset{\sim}{D}^{-1}\underset{\sim}{X})^{-1}$ form being more desirable because it obviates the assumption of homogeneity of variances. If ordinary least squares are used (e.g., should the standard errors not be available), then the residual error sum of squares in the ANOVA table must be partitioned in a one-to-one correspondence with the contrasts sums of squares to be tested in the manner of the example in Eberhart and Gardner (1966).

For case (B), in which environments are considered a random sample of possible environments, the regression model takes the form

$$E(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\alpha} + \underset{\sim}{Z}\underset{\sim}{\gamma} = \underset{\sim}{X}\underset{\sim}{\alpha} \qquad (1)$$

where E is the expectation operator and $\underset{\sim}{y}$ is as before, $\underset{\sim}{X}$ is a matrix of coefficients of the genetic parameters in $\underset{\sim}{\alpha}$ over all environments sampled, $\underset{\sim}{Z}$ is a matrix of coefficients associated with the vector $\underset{\sim}{\gamma}$, the levels of all the environmental factors: ϵ_ℓ and $\omega_{q\ell}$ for $\ell = 1, \cdots, t$. The error variance under model (1) is taken to be as before, i.e., D, estimated by the variances of the generation means and $\text{var}(\underset{\sim}{\gamma}) = \underset{\sim}{V}$ so that $\text{var}(\underset{\sim}{y}) = \underset{\sim}{Z}\underset{\sim}{V}\underset{\sim}{Z}' + \underset{\sim}{D}$. On the assumption that the environmental factors in $\underset{\sim}{\gamma}$ are mutually independent (have covariance zero), the environmental variance components in $\underset{\sim}{V}$ are estimable by Henderson's method 3 [see Searle (1971)] after an appropriate transformation on $\underset{\sim}{y}$ of (1).

Let $\underset{\sim}{L}$ be the transformation matrix such that $\underset{\sim}{LL} = \underset{\sim}{D}^{-1}$, hence $\underset{\sim}{L}$ is a diagonal matrix of the square root of the reciprocal of the variances associated with each entry and (1) becomes

$$E(\underset{\sim}{Ly}) = \underset{\sim}{LX\alpha} + \underset{\sim}{LZ\gamma}$$

or

$$E(\underset{\sim}{y^*}) = \underset{\sim}{X^*\alpha} + \underset{\sim}{Z^*\gamma} = \underset{\sim}{X^*\alpha} \ . \tag{2}$$

The $\underset{\sim}{L}$ transformation accommodates existing algorithms for the estimation of the variance components in $\underset{\sim}{V}$. The normal equations for estimation of the fixed effects in (2) by Henderson's method 3 are

$$\underset{\sim}{X^{*\prime}}(\underset{\sim}{Z^*}\hat{\underset{\sim}{V}}\underset{\sim}{Z^{*\prime}} + \underset{\sim}{D})^{-1}\underset{\sim}{X^*}\hat{\underset{\sim}{\alpha}} = \underset{\sim}{X^*}(\underset{\sim}{Z^*}\hat{\underset{\sim}{V}}\underset{\sim}{Z^{*\prime}} + \underset{\sim}{D})^{-1}\underset{\sim}{y^*} \tag{3}$$

where $\hat{V}$ is an estimate of V obtained from Henderson's method 3, for example. In order to avoid inverting $\underset{\sim}{Z^*}\hat{\underset{\sim}{V}}\underset{\sim}{Z^{*\prime}} + \underset{\sim}{D}$, we substitute for (3) Henderson's mixed model equations [Searle (1971), p. 460 or Henderson (1963)] appropriate to (2). These are

$$\begin{bmatrix} \underset{\sim}{X^{*\prime}X^*} & \underset{\sim}{X^{*\prime}Z^*} \\ \underset{\sim}{Z^{*\prime}X^*} & \underset{\sim}{Z^{*\prime}Z^*} + \underset{\sim}{V}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{\alpha}} \\ \hat{\underset{\sim}{\gamma}} \end{bmatrix} = \begin{bmatrix} X^{*\prime}y^* \\ Z^{*\prime}y^* \end{bmatrix}$$

giving

$$\hat{\underset{\sim}{\alpha}} = \left[\underset{\sim}{X^{*\prime}X^*} - \underset{\sim}{X^{*\prime}Z^*}(\underset{\sim}{Z^{*\prime}Z^*} + \underset{\sim}{V}^{-1})^{-1}\underset{\sim}{Z^{*\prime}X^*}\right]^{-1}\underset{\sim}{X^{*\prime}}\left[\underset{\sim}{I} - \underset{\sim}{Z^*}(\underset{\sim}{Z^{*\prime}Z^*} + \underset{\sim}{V}^{-1})^{-1}\underset{\sim}{Z^{*\prime}}\right]\underset{\sim}{y^*}$$

which requires inversions of matrices of lesser order than in (3). Alternately, estimates of fixed effects and of the variance components can be obtained simultaneously by the maximum likelihood (ML) procedure of Hemmerle and Hartley (1973) or also by a restricted maximum likelihood (REML) procedure [Corbeil and Searle (1974)] which for equally replicated data usually reduces to the more widely known

ANOVA procedures. One clear advantage of the ML and REML algorithms is that they guard against producing negative estimates of variances.

There are many attendant problems with estimating variance components, especially in a mixed model setting. It is expected that continuing advances in the understanding of g-e interactions will come from experiments in which environments will be controlled and varied nearly at will. The inferences made from such experiments will be restricted to these same environments but these can now appropriately be considered as fixed factors. As a consequence, the well-known weighted least squares estimation procedure yields a proper and reliable analysis.

Discussion

A lucid discussion of various genetical and statistical aspects of the present paper can be found in the 1963 symposium volume, "Statistical Genetics and Plant Breeding" [see, for example, Comstock and Moll (1963)].

The genetic model which is basic to the developments presented in this paper rests heavily on the premise that the originating varieties are in a Hardy-Weinberg equilibrium. In practice this premise is seriously violated when there exists detectable linkage disequilibrium within varieties. Departures from linkage equilibrium will also bias the derived generations in the presence of epistasis. Thus the assumption of no linkage is a crucial one. However, if the number of loci controlling a trait of interest is large compared to the number of pairs of chromosomes and the loci are randomly distributed on and among these chromosomes, each locus having only a small effect on the phenotype, then the effects of linkage will be minimal in terms of the average linkage intensity over all possible pairs of loci.

Finally, it is of interest that in practice the synthetizing method employed in obtaining varieties requires only one generation of random mating to bring the

synthetized variety to Hardy-Weinberg equilibrium in the absence of linkage. Say the pollen from line A is used exclusively to fertilize line B. In the absence of linkage the gametic array of line A will be

$$\prod_{i}^{n} \left(\sum_{j_i}^{m_i} p_{ij_ik} A_{ij_i} \right)$$

and the gametic array of line B will be

$$\prod_{i}^{n} \left(\sum_{j_i}^{m_i} p_{ij_ik'} A_{ij_i} \right) .$$

The genotypic array of variety AB will therefore be

$$\prod_{i}^{n} \left(\sum_{j_i}^{m_i} p_{ij_ik} A_{ij_i} \right) \left(\sum_{j_i}^{m_i} p_{ij_ik'} A_{ij_i} \right)$$

under random mating, and it can be shown that in turn its gametic array will be

$$\prod_{i}^{n} \left[\tfrac{1}{2} \sum_{j_i}^{m_i} (p_{ij_ik} + p_{ij_ik'}) A_{ij_i} \right]$$

and will remain such through further generations of random mating.

An Example

Several generations were derived from a cross between Red Currant tomato (Lycopersicon pimpinellifolium) and a Tangerine line (L. esculentum) for the purpose of examining the heritability of average fruit weight when locule number remained fixed at 2 or 3 per fruit. The pure-breeding parent lines, the F_1, F_2, and F_3 generations were grown simultaneously in four blocks, each with 616 plots consisting

of 56 Red Currant, 56 Tangerine, 28 F_1, 140 F_2, and 336 F_3 plants. The F_2 and F_3 plantings were divided, for the purpose of this analysis, into two and three natural groups, respectively. Each of the 2,409 observations was recorded as the average log gram weight of fruit per plant and a summary of the data is presented in Table 2.

Table 2: Fruit weight in log grams summarized by generation

	Block 1			Block 2			Block 3			Block 4		
	no.	mean	s. e.	no.	mean	s. e.	no.	mean	s. e.	no.	mean	s. e.
RC	56	0.029	.0124	56	0.056	.0077	56	0.037	.0065	56	-0.001	.0089
Tang	56	1.873	.0125	56	1.907	.0126	56	1.926	.0105	56	1.942	.0067
F_1	28	0.934	.0147	28	0.965	.0105	28	0.912	.0147	28	0.944	.0128
$F_2(1)$	70	0.956	.0225	70	1.003	.0222	70	0.973	.0189	70	0.978	.0203
$F_2(2)$	70	0.949	.0241	70	0.963	.0176	70	0.932	.0202	70	0.934	.0232
$F_3(1)$	111	0.932	.0219	111	0.947	.0225	112	0.952	.0257	112	0.931	.0268
$F_3(2)$	105	0.925	.0239	104	0.911	.0220	105	0.902	.0207	107	0.961	.0247
$F_3(3)$	101	0.860	.0226	109	0.874	.0241	108	0.894	.0239	105	0.891	.0247

RC = Red Currant
Tang = Tangerine
no. = number of plants observed
s. e. = standard error

This data is a subset of a larger experiment in which differences between blocks were observed. These differences, however, were not manifested on average fruit weight per plant as can be seen from the analysis presented in Table 3. The blocks were considered as four fixed environments and the genetic-environment interactions viewed as deviations from the best genetic model fitted over all environments simultaneously. The strategy adopted was to fit 1) the basic model; a_1^*, a_2^*, and h_{12}^*; then 2) the epistatic model by adding aa_{12} and dd_{12}; and finally 3) the g-e model by also including all genetic-environmental interactions. The procedure used is one of weighted least squares and the results are presented in Table 3.

Table 3: Estimates of genetic and environmental effects

	basic model		epistatic model		g-e model	
	estimate	s. d.	estimate	s. d.	estimate	s. d.
a_1^*	0.027	.003991	0.033	.004102	0.030	.250042
a_2^*	1.915	.004582	1.924	.004755	1.912	.250059
h_{12}^*	-0.035	.006858	0.419	.092164	0.417	.266735
aa_{12}			-0.151	.023569	-0.142	.251129
dd_{12}			-0.303	.071065	-0.307	.260114
g-e int.					$.142 > \omega_{q\ell} > -.132$, all s.d. > .25	
d. f.	29		27		12	
$\chi^2_{\alpha=.05}$	5.07 n. s.		3.23 n. s.		2.06 n. s.	

Sequential reduction in sums of squares

		S.S.	M.S.	F-test
Total	32	221,094.44		
Basic model	3	220,947.54	73,649.18	sig.
Epistatic model	2	59.80	29.90	sig.
g-e model	15	62.40	4.16	n. s.
Residual	12	24.70	2.06	n. s.

sig. = probability $\leq .05$

n. s. = probability $> .05$

$\hat{\omega}_{q\ell}$ = deviation in q^{th} genetic parameter due to ℓ^{th} environment (block)

s. d. = standard deviation of the estimate

It is noted that the basic model adequately explains the variation observed between entries but that a further significant reduction in the residual sums of squares can be achieved by fitting the epistatic components. An examination of the estimates obtained under the epistatic model reveals each to be quite different from zero, whereas this conclusion is not justified under the g-e model. Therefore, it is concluded that epistasis plays an important role in average fruit weight per plant in the inter-specific cross under study. Further, the heterosis estimate h^*_{12}, can reasonably be expected to contain some epistatic effects confounded with dominance deviation. Finally, the difference between a^*_1 and a^*_2 leaves no doubt as to the high heritability which exists in average fruit weight. However, it is not quite so clear what portion of this difference can be attributed to additive gene effects and what portion is due to the additive x additive gene interaction. A more detailed genetic analysis of this experiment is possible but it is not in keeping with the present objectives, but rather we elucidate some of the details of the analysis.

The epistatic model assumes that each entry consists of a linear combination of the genetic effects a^*_1, a^*_2, h^*_{12}, aa_{12}, and dd_{12}. These combinations are available from Table 1. If we let RC be variety 1 and Tangerine be variety 2, then including the F_1's, F_2's, and F_3's we have the following model equations:

$$\begin{aligned}
&\text{RC:} && Y_1 = a^*_1 \\
&\text{Tang:} && Y_2 = a^*_2 \\
&F_1: && Y_{12} = \tfrac{1}{2}a^*_1 + \tfrac{1}{2}a^*_2 + h^*_{12} + aa_{12} + dd_{12} \\
&F_2: && Y^2_{12} = \tfrac{1}{2}a^*_1 + \tfrac{1}{2}a^*_2 + \tfrac{1}{2}h^*_{12} + aa_{12} + \tfrac{1}{4}dd_{12} \\
&F_3: && Y^3_{12} = \tfrac{1}{2}a^*_1 + \tfrac{1}{2}a^*_2 + \tfrac{1}{4}h^*_{12} + aa_{12} + \tfrac{1}{16}dd_{12} .
\end{aligned}$$

The number of equations is expanded in two ways: 1) recall that two independent observations were made on the F_2 generation and three were made on the F_3

generation, and 2) the eight equations implied in 1) above are applied to each of the four blocks. In an abbreviated matrix form, this gives

$$\begin{bmatrix} Y_{1.1} \\ Y_{2.1} \\ Y_{12.1} \\ Y^2_{12.1} \\ Y^2_{12.1} \\ Y^3_{12.1} \\ Y^3_{12.1} \\ Y^3_{12.1} \\ Y_{1.2} \\ \vdots \\ Y^3_{12.4} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 1 & 1 & 1 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1 & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1 & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & 1 & \frac{1}{16} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & 1 & \frac{1}{16} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & 1 & \frac{1}{16} \\ 1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & 1 & \frac{1}{16} \end{bmatrix} \begin{bmatrix} a^*_1 \\ a^*_2 \\ h^*_{12} \\ aa_{12} \\ dd_{12} \end{bmatrix}$$

or, more briefly, $E\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta}$.

Now replacing $E\underset{\sim}{y}$ by the observed generation means, $\underset{\sim}{y}$, and multiplying the first element of $\underset{\sim}{y}$ and each element of the first row of $\underset{\sim}{X}$ by $1/.0124$, the second element of $\underset{\sim}{y}$ and each element of the second row of $\underset{\sim}{X}$ by $1/.0125$, etc. where the multiplier is taken as the reciprocal of the s. e. values from Table 2 yields in matrix notation

$$\underset{\sim}{y}^* = \underset{\sim}{X}^*\underset{\sim}{\beta} + \underset{\sim}{\epsilon}^*$$

and the weighted least square estimates of $\underset{\sim}{\beta}$, $\hat{\underset{\sim}{\beta}}$ say, are $\hat{\underset{\sim}{\beta}} = (\underset{\sim}{X}^{*\prime}\underset{\sim}{X}^*)^{-1}\underset{\sim}{X}^{*\prime}\underset{\sim}{y}$ with $\text{var}(\hat{\underset{\sim}{\beta}}) = (\underset{\sim}{X}^{*\prime}\underset{\sim}{X}^*)^{-1}$ and residual $= \underset{\sim}{y}^{*\prime}\underset{\sim}{y}^* - \underset{\sim}{y}^{*\prime}\underset{\sim}{X}^*\hat{\underset{\sim}{\beta}}$ which is distributed as χ^2. The term $\underset{\sim}{y}^{*\prime}\underset{\sim}{y}^*$ is the total sum of squares and $\underset{\sim}{y}^{*\prime}\underset{\sim}{X}^*\hat{\underset{\sim}{\beta}}$ is the regression sum of squares for fitting the epistatic model.

On fitting the basic model the matrices and procedures are identical except that the 4^{th} and 5^{th} columns of $\underset{\sim}{X}$ do not appear nor does aa_{12} and dd_{12} appear in the vector $\underset{\sim}{\beta}$.

The $\underset{\sim}{X}$ matrix and vectors associated with the g-e model have the form

$$\begin{bmatrix} \underset{\sim}{Y}_1 \\ \underset{\sim}{Y}_2 \\ \underset{\sim}{Y}_3 \\ \underset{\sim}{Y}_4 \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}_0 & \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{X}_0 \\ \underset{\sim}{0} & \underset{\sim}{X}_0 & \underset{\sim}{0} & \underset{\sim}{X}_0 \\ \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{X}_0 & \underset{\sim}{X}_0 \\ -\underset{\sim}{X}_0 & -\underset{\sim}{X}_0 & -\underset{\sim}{X}_0 & \underset{\sim}{X}_0 \end{bmatrix} \begin{bmatrix} \underset{\sim}{\omega}_1 \\ \underset{\sim}{\omega}_2 \\ \underset{\sim}{\omega}_3 \\ \underset{\sim}{\beta} \end{bmatrix}.$$

Suppose block 4 is a standard or reference environment, then

$\underset{\sim}{Y}_1$, e.g., is the vector of expected generation means in block 1;

$\underset{\sim}{\omega}'_\ell = (\omega_{a_1^* \ell-4}, \omega_{a_2^* \ell-4}, \omega_{h_{12}^* \ell-4}, \omega_{aa_{12} \ell-4}, \omega_{dd_{12} \ell-4})'$, where $\ell = 1, 2, 3$

and $\omega_{a_1^* \ell-4}$ means the interaction of a_1^* with block ℓ contrasted with the interaction of a_1^* with block 4, and so on with the other genetic-environmental interaction;

$\underset{\sim}{X}_0$ is a matrix consisting of the first eight rows of $\underset{\sim}{X}$ as defined for the epistatic model;

and

$\underset{\sim}{0}$ is an 8 by 5 matrix of all zeros.

References

Comstock, R. E. and R. H. Moll (1963). Genotype-environment interactions. In Statistical Genetics and Plant Breeding, NAS-NRC Publication 982.

Corbeil, R. R. and S. R. Searle (1974). Restricted Maximum Likelihood (REML) estimation of variance components in the mixed model. BU-538-M, Biometrics Unit, Cornell University, Ithaca, New York.

Eberhart, S. A. and C. O. Gardner (1966). A general model for genetic effects. Biometrics 22:864-881.

Eberhart, S. A. and W. A. Russell (1966). Stability parameters for comparing varieties. Crop Sci. 6:36-40.

Fisher, R. A. and W. A. Mackenzie (1923). Studies in Crop Variation. II. The manurial response of different potato varieties. J. Agric. Sci., Camb. 13:311-320.

Freeman, G. H. (1973). Statistical methods for the analysis of genotype-environment interactions. Heredity 31:339-354.

Hayman, B. I. (1954). The analysis of variance of diallel tables. Biometrics 10: 235-244.

Hemmerle, W. J. and H. O. Hartley (1973). Computing maximum likelihood estimates for the mixed A.O.V. model using the W-transformation. Technometrics 15: 819-832.

Henderson, C. R. (1963). Selection index and expected genetic advance. In Statistical Genetics and Plant Breeding, NAS-NRC Publication 982.

Kempthorne, O. (1957). An Introduction to Genetic Statistics. Wiley and Sons, Inc., New York.

Knight, R. (1973). The relation between hybrid vigour and genotype-environment interactions. Theor. and Appl. Genetics 43:311-318.

Searle, S. R. (1971). Linear Models. Wiley and Sons, Inc., New York.